ChinaXiv合作期刊

基于链路预测的有向互动影响力和用户信任的推荐算法 *

魏映婷，倪 静+

(上海理工大学 管理学院，上海 200093)

摘 要：针对传统推荐算法存在忽视社交网络结构紧密强度对用户信任传递的影响和缺乏社交心理解释等问题，提出基于链路预测的有向性互动影响力和用户信任的推荐算法。首先利用融合用户偏好行为和社交圈的综合相似度识别出目标用户的相似朋友圈；其次通过结合节点引力指数和有向性影响因子获得目标用户之间的有向性互动影响力，再利用由有向性互动影响力和用户评分信任而得的综合用户信任值在目标用户的相似朋友圈中寻找出值得信任的相似用户集合，这有效提高了推荐的精确性，最后产生推荐。结果表明，所提的推荐方法较之前的社会网络推荐算法在性能上具有显著提高。

关键词：社交网络；评分信任度；有向性互动影响力；链路预测

中图分类号：TP doi: 10.19734/j.issn.1001-3695.2018.10.0801

Recommendation algorithm based on link prediction for directed interaction influence and user trust

Wei Yingting, Ni Jing+

(*Business School*, *University of Shanghai for Science & Technology*, *Shanghai* 200093, *China*)

Abstract: Aiming at the problems that the traditional recommendation algorithm ignores the influence of the tight structure of social network structure on user trust transmission and the lack of social psychological explanation, a recommendation algorithm based on link prediction for directed interaction and user trust is proposed. Firstly, the similar user circle of the target user is identified by the integrated similarity between the user preference behavior and the social circle. Secondly, the directional interaction influence between the target users is obtained by combining the node gravity index and the directed influence factor. The integrated user trust value of the directional interaction influence and the user score trust finds a trustworthy similar user set in the similar circle of friends of the target user, which effectively improves the accuracy of the recommendation and finally generates the recommendation. The results show that the proposed method has a significant improvement in performance compared to the previous social network recommendation algorithm.

Key words: social network; score trust; directive interaction; link prediction

chinaXiv:201904.00049v1

0 引言

目前，已有不少学者将信任引入电子商务推荐系统，从而缓解了推荐准确性、数据稀疏度等问题，李良等人[1]提出将评分信任度和偏好信任度线性加权融合，改进信任度计算方法，提高了推荐的准确率。丁小焕等人[2]提出了一种融合朋友关系和标签信息的张量分解推荐算法，该算法验证了结合朋友关系能够提高推荐的准确率。Zhu 等人[3]根据原有信任关系网络，结合不同的信任传播方式，挖掘出未知的信任关系，但该方法只借助于用户已经记录的信任网络信息，而忽略了用户间的交互信息，Liben-Nowell 等人[4]将相似性指标分为基于节点和基于路径的指标，并分析了若干指标合作对社会网络中链接预测的效果。

推荐系统与社交网络的融合使得个性化推荐发展到一个新阶段。众所周知，人们一般比较容易接受朋友和专家的推荐。所以考虑将用户信任和互动影响力融入推荐算法中解释了用户社交心理对用户社交行为的影响。该推荐算法把融合有向互动影响力的用户信任作为影响推荐的重要因素：a) 本文结合 RA 指数 (resource allocation index) [5]和 AA 指数 (adamic-adar index) [6]，提出了利用一种节点互动影响预测方法—有向性节点引力指数 ，有向性节点引力指数是在考虑被预测两个节点的共同邻居节点的节点度的基础上,同时考虑被预测节点与共同邻居组成的小型网络对预测用户互动影响力的影响,由于每个用户的兴趣点不同，愿意接受的其他用户对自身的影响也不同，所以本文还应考虑互动影响力的有向性[7]；b)另一方面，考虑整个推荐系统中用户信任度，从用户活跃度和用户评分客观度两个因素来计算用户信任度。本文的主要的推荐过程是首先根据融合偏好行为和社交圈的用户相似度找到目标用户的相似邻居集，然后在相似邻居集中选择推荐信任值较高的 top-k 用户推荐给目标用户。

1 相关研究

1.1 典型的链路预测方法

在复杂网络链路预测的研究中，网络链路的预测方法已经成为研究热点，链路预测是指如何通过已知的网络节点以及网络结构等信息，预测网络中尚未产生连边的两个节点之间产生连接的可能性[8]。典型的 RA 指数和 AA 指数计算节点相似度时将共同邻居节点度信息考虑进来，分别取得较好的预测效果。RA 指数[5]：该算法从资源分配的角度出发，认为网络中每个节点都有一定的资源，并将资源平均分配给它的邻居。网络中没有直接相连的两个节点 x,y 可以从节点 x 传递一些资源到节点 y，在此过程，节点 x 和 y 的共同邻居

收稿日期：2018-10-23；**修回日期：**2018-12-04 **基金项目：**国家自然科学基金资助项目（7177411）

作者简介：魏映婷（1994-），女，湖南衡阳人，硕士研究生，主要研究方法为信息管理系统、决策支持系统、复杂网络；倪静（1972-），女（通信作者），副教授，博士，主要研究方向为管理信息系统、电子商务、复杂网络（nijing501@126.com）.

ChinaXiv合作期刊

成为传递媒介，则节点 y 接收到的资源数就定义为节点 x,y 的相似度记为 RA 指数。AA 指数[6]：该算法的思想是节点度小的共同邻居节点对于相似性的作用程度反而大于节点度大的共同邻居节点，基于该思想定义节点之间的节点相似性记为 AA 指数。

这两种指数只考虑被预测的两个节点共同邻居数目和共同邻居的节点度数，并未深入研究预测用户与它们的共同邻居之间所形成的小型网络结构紧密程度以及节点出入度的方向性对预测结果的影响。

1.2　好友信任的度量

信任是通过收集和分析用户之间的互动行为获得的用于衡量好友关系密切程度的度量。之前的学者研究用户信任一般考虑用户评分认可度和用户评分准确度，用户评分认可度表示用户对项目的评分被大众认可的程度，用户评分准确度表示用户的评分和大众对某项目评分平均值的接近程度。他们都是简单的从这两个因素考量用户信任，这样的方法比较片面，使得推荐不够准确，没有挖掘出潜在的信任好友。

1.3　协同过滤算法与相似度计算

传统的协同过滤推荐算法过程：a)根据用户—项目评分矩阵计算用户相似度；b)根据相似度选取近邻用户；c)根据近邻用户给出预测分数；d)产生推荐。本文的相似度计算如下：

a) 基于偏好行为的用户相似度。

本文根据 IF-IDF 算法[9]的思路计算项目标签对于用户爱好物体的权重 $\mathrm{w}(u_i,b_{t_i})$，计算公式如式（1）所示。

$$\mathrm{w}(u_i,b_{t_i})=TF_{(u_i,b_{t_i})}\times IDF_{b_{t_i}} \tag{1}$$

其中：$TF_{(u_i,b_{t_i})}$ 表示标签 b_{t_i} 在用户评分项目中出现的频率；$IDF_{b_{t_i}}$ 表示该标签在所有用户的评分项目模型中的区别能力，它们的计算方法分别表示为

$$TF_{(u_i,b_{t_i})}=\frac{num_{(u_i,b_{t_i})}}{num_{(u_i)}} \tag{2}$$

其中：$num_{(u_i,b_{t_i})}$ 表示用户 u_i 使用标签 b_{t_i} 的次数；$num_{(u_i)}$ 表示用户 u_i 使用标签的总数。

$$IDF_{(b_{t_i})}=log\frac{num_{user}}{num_{t_i}} \tag{3}$$

其中：num_{user} 为推荐系统中的用户总数；num_{t_i} 为使用过标签 t_i 的用户的数量。综上所述，可得出基于标签系统的项目权重向量($\mathrm{w}_{\mathrm{item}_i}$)为如式(4)所示。

$$\mathrm{w}_{\mathrm{item}_i}=(\mathrm{w}(u_i,b_{t_1}),\mathrm{w}(u_i,b_{t_2}),\ldots,\mathrm{w}(u_i,b_{t_i})) \tag{4}$$

其中：item_i 代表用户 i 对项目 item 的偏好行为。综上，基于偏好行为的用户相似度($\mathrm{sim}^{\mathrm{tag}}_{(u_i,u_j)}$)为

$$\mathrm{sim}^{\mathrm{tag}}_{(u_i,u_j)}=\cos(u_i,u_j)=\frac{\sum_{i,j\in(I_{u_i}\cap I_{u_j})}\mathrm{w}_{item_i}\cdot\mathrm{w}_{item_j}}{\sqrt{\sum_{i\in(I_{u_i}\cap I_{u_j})}\mathrm{w}_{item_i}^{2}}\sqrt{\sum_{j\in(I_{u_i}\cap I_{u_j})}\mathrm{w}_{item_j}^{2}}} \tag{5}$$

其中：$I_{u_i}\cap I_{u_j}$ 表示用户 u_i 和 u_j 的共同评分项目集合。

b) 用户社交圈的相似度。

在社交网路中，如果两个用户拥有重叠程度很高的社交圈，那他们很可能会建立朋友关系。为了利用社交圈发现潜在朋友，首先需要识别社交圈，即实现对朋友的自动分组。如果两个用户之间“朋友圈”交集集合越大，这两用户之间兴趣爱好生活方式越相似。文中通过计算两个用户的共同朋友数目来确定用户社交圈的相似性($sim^{social}_{(u_i,u_j)}$)为

$$sim^{social}_{(u_i,u_j)}=\frac{|friend_{u_i}\cap friend_{u_j}|}{|friend_{u_i}\cup friend_{u_j}|} \tag{6}$$

其中：$friend_{u_i}$ 和 $friend_{u_j}$ 表示用户 u_i 和 u_j 的朋友集合。

2　基于链路预测的有向性互动影响力和用户信任的推荐算法

RUTI-CF 算法分为两个具有递进性的推荐过程，首先以目标用户进行聚类分析形成以目标用户为中心的相似用户集合，再从相似用户集合中寻找值得信任的相似用户进行推荐。算法模型如图 1 所示。

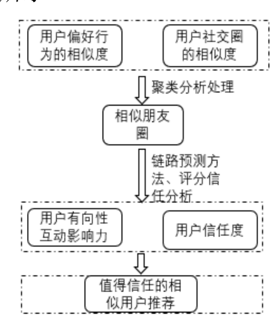


图 1　RUTI-CF 算法模型

Fig. 1　RUTI-CF algorithm model

2.1　融合偏好行为和社交圈的用户相似度

在现实生活中，人们往往会更愿意接受朋友圈中好友的推荐以及更爱和自己个性爱好相似的推荐。相似度计算的改进优化是推荐算法的关键所在。随着社交网络的不断发展，交互用户在扮演者越来越活跃的角色，本文在考量用户相似度的计算不限于平常的购买、评分和标签行为，还会考虑朋友圈中喜爱偏好相投的其他线上虚拟好友用户的偏好行为。用户的抉择在一定程度上会受到这些用户的影响。

基于上述原因，本节提出了一种结合用户偏好行为和社交圈的相似性度量。一方面由于用户—标签数据、项目—标签的稀疏性，决定了用户相似度也会出现稀疏性的缺陷；另一方面从基于用户社交圈的相似性分析考量同一个社交圈内的用户之间的联系更为紧密，在社交网络中，用户的偏好行为和社交圈对计算用户相似度的贡献都是不可小觑。综上相关研究，当基于用户偏好行为的相似值($*sim^{tag}_{(u_i,u_j)}$)和基于社交圈的相似值($sim^{tag}_{(u_i,u_j)}$)都存在时，本文则用调和平均值去融合两者；反之，则取两个相似值的最大值。这样有效地缓解了用户相似矩阵的数据稀疏性。融合计算如式(7)所示。

$$\mathrm{sim}_{(u_i,u_j)}=\begin{cases}\dfrac{2*\mathrm{sim}^{\mathrm{tag}}_{(u_i,u_j)}*\mathrm{sim}^{\mathrm{social}}_{(u_i,u_j)}}{\mathrm{sim}^{\mathrm{tag}}_{(u_i,u_j)}+\mathrm{sim}^{\mathrm{social}}_{(u_i,u_j)}} & \mathrm{sim}^{\mathrm{tag}}_{(u_i,u_j)}>0\text{且}\mathrm{sim}^{\mathrm{social}}_{(u_i,u_j)}>0\\ \max\{\mathrm{sim}^{\mathrm{tag}}(u_i,u_j),\mathrm{sim}^{\mathrm{social}}(u_i,u_j)\} & \text{其他情况}\end{cases} \tag{7}$$

基于以上的用户相似度计算方法本文可以得到社交网络中每个用户的以自我为中心的小型相似用户网络结构圈。下面本文将用综合用户评分行为的信任关系和用户节点影响力的推荐信任衡量这个圈中的每个相似用户的推荐信任值。

2.2　综合有向性互动影响力和用户信任的推荐信任度

2.2.1 用户信任度

1)用户活跃度

chinaXiv:201904.00049v1

chinaXiv:201904.00049v1

推荐系统中用户活跃度[10]存在较大的差异，用户活跃度可以用于隐性推断用户对未评分项目感兴趣的可能性。用户的活跃度越高，用户对未评价项目不感兴趣的可能性越大。所以本文引入用户活跃度可以作为衡量用户评分行为可信度的一个因子。本文提出的用户活跃度($\text{Activity}_{u,item}$)定义为

$$\text{Activity}_{u,item}=\frac{\sum n_{u,item}}{n_{u,item}}\log\frac{user}{user_{item}} \tag{8}$$

其中：$n_{u,item}$ 为用户 u 评分某一项目 *item* 的次数；*user* 表示用户总数；$user_{item}$ 表示评分项目 *item* 的用户数目。所以，用户 *u* 评分项目 *item* 的次数越多，表示他在这一领域的活跃度越高，权威性也会越大，同时若用户的好友评分项目次数越多，代表此用户在此项目的评论影响力越大，传播给其他好友的影响较大，同时说明用户对该项目所属领域越了解，评分的可信度越高;反之则评分可信度越低。

2)用户 *u* 对项目 *item* 的评分客观度

一个用户的信任程度比较高，只是表明这个用户值得信任，他的评分整体可信度较高而已，并不代表用户每次评分都那么可信和客观，这个问题反映在表现为用户关于单个项目评价的可信性。考虑到用户打分时的随意性也会影响评分的客观公正，所以用户 u 的评分客观度($Objectivity_{u,item}$)定义如下:

$$Objectivity_{u,item}=1-\frac{\left|r_{u,item}-\overline{r_{item}}\right|}{r_{max,item}-r_{min,item}} \tag{9}$$

其中：$r_{u,item}$ 表示用户 u 对项目 item 的评分值；$\overline{r_{item}}$ 则表示社交网络中所有的用户对此项目 item 的评分平均值；$r_{max,item}$ 代表所有的用户对此项目评分数据中的最高评分值；$r_{min,item}$ 表示所有的用户对项目评分数据中的最低评分值。用户 u 关于项目的评价越接近评分均值，则用户的评分客观度越接近 1。$Objectivity_{u,item}$ 取值在 0~1 之间。

由用户活跃度和用户评分客观度可以计算用户 u 在某个领域的评分准确度，用户 *u* 在项目 *item* 的活跃度越大，评分越客观，那么他在项目 item 的评分价值性越高，他在这个领域越值得被其他用户信任。所以用户 *u* 在项目 *item* 上的评分可信度($\text{Scored_Value}_{u,item}$)定义为

$$\text{Scored_Value}_{u,item}=\text{Activity}_{u,item}*Objectivity_{u,item} \tag{10}$$

得出被推荐用户 u_j 对于目标用户 u_i 之间的信任度($Trust^{rating}(u_i,u_j)$)为

$$Trust^{rating}(u_i,u_j)=\frac{\sum_{item\in(I_{u_i}\cap I_{u_j})}\text{Scored_Value}_{u_j,item}}{\left|I_{u_i}\cap I_{u_j}\right|} \tag{11}$$

其中：$I_{u_i}\cap I_{u_j}$ 表示用户 u_i 和 u_j 的共同评分项目集合。

2.2.2 基于链路预测的有向性用户互动影响力

在社交网络中，本文可以将有共同项目的用户认为是邻居节点，则两个邻居节点的度 $d_{u_iu_j}$ 定义为

$$d_{u_iu_j}=\begin{cases}1?I_{u_i}\cap I_{u_j}\neq\varnothing\\0?I_{u_i}\cap I_{u_j}=\varnothing\end{cases} \tag{12}$$

其中：I_{u_i} 和 I_{u_j} 分别用户 u_i,u_j 的项目集合，表示当 $I_{u_i}\cap I_{u_j}\neq\varnothing$ 表示用户 u_i,u_j 有共同评分项目，那么会存在一条从节点 u_i 指向 u_j 的有向边时，它们之间的节点度 $d_{u_iu_j}=1$，反之两个节点的节点度 $d_{u_iu_j}=0$。

用有向带权图 G = (V,E,I) 作为社会网络的抽象表述，其中：V 代表网络中的所有节点集合，用户之间的具体关系则用每节点之间的连边表示；E 代表边的集合。在网络图形中，边权值 I_{ij} 表示节点 i 指向节点 j 的互动影响值。给定社交网络 G，本文的目标是通过对社交网络中各好友之间互动影响力的分析来预测尚未建立信任关系的两个人之间建立信任关系的可能性。所以用户之间的互动影响关系如图 2 所示。

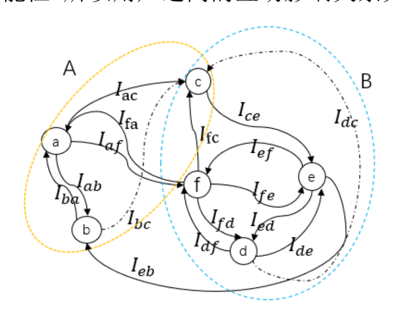


图 2　互动影响力图

Fig. 2　Interactive influence diagram

如图 2 所示，本文可以得出节点 a 传播至节点 b 的互动影响力 I_{ab} 与节点 b 传播至节点 a 的互动影响力 I_{ba} 是不平衡的，体现了互动影响力的有向性。同时，节点 a 对节点 c 产生了单向的互动影响力 I_{ac}，由于互动影响力的传递性，本文可以得知节点 b 会对节点 c 产生一定的互动影响力 I_{bc}。本文可以计算 I_{bc} 来预测节点 b 至 c 产生信任链接的可能性；节点间的互动影响力在会增强节点间的信任关系的建立。

本文把用户节点的评分项目类别看做它的兴趣偏好点，这样用户节点的共同评分项目类别可以当作用户之间的相同兴趣偏好点，目标用户可以通过相同兴趣偏好点开辟一条兴趣偏好路径进而影响与其可能产生信任关系的用户；所以本文可以计算共同评分项目个数所占比例来确定有向性影响因子，以表现用户之间相互影响的差异性。当目标用户与其可能产生信任关系的用户拥有越多的共同评分项目，而且共同评分项目占目标用户的评分项目总数目的比例越高，说明目标用户有多条兴趣偏好路径倾向于对此用户产生影响，反之产生的影响因子越低。有向性影响因子($\gamma_{(u_i,u_j)}$)的定义公式为

$$\gamma_{(u_i,u_j)}=\frac{\left|I_{u_i}\cap I_{u_j}\right|}{\left|I_{u_i}\right|} \tag{13}$$

其中：$I_{u_i}\cap I_{u_j}$ 表示用户 u_i 和 u_j 的共同评分项目集合。

本文提出的互动影响力考量节点 b 和 c 产生信任传递链接的可能性时不仅考虑被预测两个节点之间的共同邻居的节点度数，同时还要考虑了被预测节点与共同邻居组成的小型网络结构的紧密程度。所以本文利用节点引力指数($S_{(u_i,u_j)}$)来预测节点 u_i 和 u_j 的互动影响力，它的公式定义为

$$S_{(u_i,u_j)}=\sum_{a\in\varphi(u_i)\cap\varphi(u_j)}{}^{e_a}\!/\!_{k_a} \tag{14}$$

$$e_a=\sum_{u_i\notin C}d_{au_i} \tag{15}$$

$$k_a=\sum_{u_i\notin D}d_{au_i} \tag{16}$$

其中：e_a 为节点 a 与其他共同邻居之间及与节点 u_i 和 u_j 的链接数；d_{au_i} 为节点 a 至节点 u_i 的节点度；k_a 是节点 u_i 和 u_j 的

ChinaXiv合作期刊

共同邻居节点的节点度；C 为节点 u_i 和 u_j 以及它们之间的共同邻居组成的小型网络用户节点集合；D 为整个社交网络中所有用户节点集合。

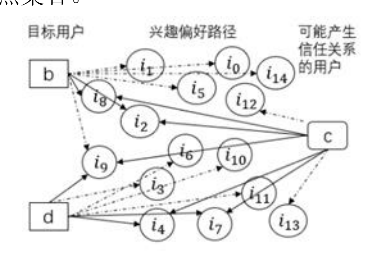


图 3　有向性影响系数示意图点

Fig. 3　Schematic point of directional influence coefficient

数。因为考虑到用户节点相互影响的差异性，本文对影响力度量方法进行修正，所以有向性的互动影响力($I^{Dir}{}_{(u_i,u_j)}$)定义如下:

$$I^{Dir}{}_{(u_i,u_j)} = \gamma_{(u_i,u_j)} * S_{(u_i,u_j)} \tag{17}$$

其中：$S_{(u_i,u_j)}$ 是节点 u_i 和 u_j 之间的互动影响力；$\gamma_{(u_i,u_j)}$ 是有向影响系数；两者的乘积得到有向性的互动影响力即社交网络 G 中的节点 b 对 c 产生的有向边权值 I_{bc} （虚线表示)。

如图 2 所示，在由节点 a、b、c 组成的小型社交网络 A 和由节点 c、d、e、f 的小型社交网路 B 中，本文用式(14)式计算节点 b 对节点 c 的产生信任链接的可能性大小为:2/3(a)=2/3，同样的方法得到节点 d 对节点 c 的产生信任链接的可能性大小为:3/4(f)+2/3(e)=17/12，节点 b、c、d 的交互行为如图 3 所示。根据式(13)本文可以得到节点 b 到节点 c 的有向性影响因子为 2/7，节点 d 到节点 c 的有向性影响因子为 3/7，最后本文根据式(17)可得节点 b 对节点 c 的有向性的互动影响力为：2/3 x 2/7=4/21；同时得到节点 d 到节点 c 的节点 d 到节点 c 的：17/12 x 3/7=17/28。所以当社交网络 B 中节点交互行为比社交网络 A 更为密切，从而社交网络 B 的结构比 A 更为紧密，节点 d 对节点 c 相比于节点 b 对节点 c 更容易建立信任关系。

2.2.3 综合推荐信任度

综合前文研究，提出融合在线社交网络用户信任($Trust^{rating}(u_i,u_j)$)和互动影响力($I^{Dir}{}_{(u_i,u_j)}$)的推荐信任度($trust_{(u_i,u_j)}$)计算方法如下:

$$trust_{(u_i,u_j)} = \alpha Trust^{rating}(u_i,u_j) + (1-\alpha) I^{Dir}{}_{(u_i,u_j)} \tag{18}$$

其中：$0 \le \alpha \le 1$，用于权衡用户信任和互动影响力对推荐信任值的影响程度，α 数值具有不确定性，将视社交网络的具体情况而定。

3　基于链路预测的有向性互动影响力和用户信任的推荐

本文按照融合偏好行为和社交圈的相似值计算方法求得用户相似值，利用用户相似值找出。用户的“朋友圈”，获取目标用户的近邻集合 NB。采用如下方法产生推荐的结果：首先假设目标用户 u 的“朋友圈”组成的集合为 NB，对于任一 $v_i \in NB$，计算目标用户 u 与朋友圈内好友 v_i 之间融合有向性互动影响力和用户信任的综合推荐信任值，利用 top-k 法获取目标用户的值得信任的近邻集合为：$TNB = \{u_1, u_2, \ldots, u_k\}$，对于任一 $u_i \in TNB$，获得其历史评价的项目集合 I_i (i=1,2,…,k)组成新的集合，将这个新的项目集合 Recommend-Item 推荐给用户。

4　实验结果及分析

4.1　实验数据集

本实验采用的实验数据是 *Last.fm* 数据集。实验评估前，针对 *Last.fm* 数据集，要求每个用户至少给 10 个音乐家标注过标签。经过筛选最后保留 881 个用户及其所有的好友信息记录和标签行为记录。其次，将这两组数据都划分为训练集和测试集两部分。本文使用交叉验证法共进行五次实验，得到五组实验结果，然后求出平均值作为最后的结果。

4.2　实验评估指标

准确率(precision)、召回率(recall)和 F1-measure 是推荐系统中常用的三个重要指标。准确率表征的是项目物体被成功推荐的比例,准确率的值越大，推荐模型越好;召回率则指的是推荐出的好友数量，占测试数据当中的好友数量的比例，召回率的值越大，模型效果越好；F1-measure 指标则是综合准确率和召回率这两个指标衡量，当两个指标都是越高越好，但是在一些情况下两者会有矛盾，此时就需要 F1-measure 指标来进行衡量，当 F1-measure 指标具有较高的值时，说明此算法的推荐效果更优。计算公式如下：

$$\text{precision} = \frac{n_{like}}{n_{recommend}} \tag{19}$$

$$\text{recall} = \frac{n_{like}}{n_{test}} \tag{20}$$

$$\text{F1-measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{21}$$

其中：$n_{recommend}$ 表示给推荐项目集合的元素个数；n_{test} 表示测试集中物体的数量；n_{like} 表示在推荐的项目中受该用户喜欢的项目总数。

4.3　实验结果分析

本实验需对 α 和 K 的合理取值进行验证，然后在合理的 α 值使得实验结果最优的情况下，将本文提出的算法与其他推荐算法进行性能比较，得出结论。

1)权重值 α 对算法的影响

分别将 α 的取值为 0.0,0.1,0.2,…,1.0 进行实验。为了消除邻居用户数量对推荐结果的影响，实验过程中依次将邻居用户的数量设为 2,3,4,…,10,进行 9 组实验。表 1 表示在 *Last.fm* 数据集上，不同的 α 对 RUTI-CF 算法的准确率的影响。可以看出不同的 α 值对本文基于链路预测的有向性互动影响力和用户信任的推荐算法(RUTI-CF 算法)准确率的影响。从表 1 中可以看出，当 K 一定时，随着 α 从 0.0 开始不断增加，RUTI-CF 算法准确率波动的趋势总体上是一致的，都是呈现先平缓上升，到达最高点，

表 1　不同 α 对 RUTI-CF 算法的准确率的影响

Table 1　Effect of different α on accuracy of RUTI-CF algorithm

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Top-2	0.1551	0.1604	0.1610	0.1622	0.1632	0.1698	0.1723	0.1621	0.1597	0.1504	0.1109
Top-3	0.2201	0.2291	0.2301	0.2434	0.2311	0.2506	0.2557	0.2272	0.2237	0.2117	0.1817
Top-4	0.3433	0.3512	0.3523	0.3567	0.3588	0.3599	0.3669	0.3315	0.3206	0.3008	0.1653
Top-5	0.4018	0.4122	0.4616	0.4656	0.4717	0.4837	0.5272	0.4589	0.4789	0.5537	0.2966
Top-6	0.4218	0.5796	0.6012	0.6082	0.6271	0.6376	0.6579	0.5989	0.6366	0.6366	0.4704
Top-7	0.5154	0.4276	0.4204	0.4213	0.4301	0.4321	0.4381	0.4112	0.4019	0.3904	0.2148
Top-8	0.3872	0.3889	0.3981	0.3915	0.3959	0.4008	0.4035	0.3779	0.3743	0.3717	0.2306
Top-9	0.3109	0.3111	0.3210	0.3219	0.3267	0.3317	0.3549	0.3093	0.2945	0.2857	0.1633
Top-10	0.2690	0.2696	0.2807	0.2817	0.2877	0.2899	0.3064	0.2754	0.2708	0.2655	0.1521

chinaXiv:201904.00049v1

ChinaXiv合作期刊

再平缓下降的趋势。当 α 取值为 0.6 时，推荐模型的准确率的值是最好。即在本文的推荐算法中，用户信任和用户之间的互动影响力对最后的推荐结果的影响比例是 3:2。说明用户在进行项目选择时，在相似用户集成的"朋友圈"中信任好友的评分对其的影响还是比用户之间的互动影响力的影响程度要大，当 α 在 0.8~1.0 内变化时，推荐准确率急剧下降。特别是当 α=1.0 时，即只考虑用户节点影响力时，与 α=0.9 时相比较，推荐推荐率下降 10%~20%，这说明不考虑互动影响力，单独考虑用户信任时，推荐准确率将大大下降，这表明用户之间的互动影响力对于用户建立信任关系有很大的促进加强作用。

2)Top-K 取值对算法的影响

根据本文产生推荐的方法，将 K 的取值设为 2,3,4,…,10，实验观察推荐准确率随着 K 值不断变化的规律。同样从表 1 的实验结果数据得知，不管权重值 α 的取值，随着 K 值的不断增加，RUTI-CF 算法准确率呈现先上升，等达到最高峰，再下降的趋势。

在本文实验环境下，当 K 的值在 2~5 内变化时，算法准确率不断上升；当 K=6 时，准确率值最高；当 K 的值在 6~10 区间内变化时，推荐的准确率又开始下降。因此，实验说明邻居数量的大小能够推荐的质量，但不是邻居数量越多，推荐的质量越好。对于本实验的数据环境而言，当目标用户的近邻取值为 6 时，实验结果最好。

3)不同算法的推荐效果比较

为了验证本文提出推荐算法的有效性，将 RUTI-CF 算法与其他算法进行对比分析，分别是 FT-CF 算法 (hybrid collaborative filtering recommendation algorithm based on friendships and tags,基于好友关系和标签的混合协同过滤算法) [11]、FRSN-CF 算法 (research on friend recommendation methods in social networks，社交网络中的好友推荐方法研究) [12]、Trust_CF 算法代表本文只考虑在线社交网络用户信任的算法、Influence_CF 算法代表本文只考虑用户之间互动影响力的推荐算法。通过实验，得到以上方法的准确率值和 F1-measure 指标值，分别如图 4 和 5 所示。

如图 4 和 5 所示，随着 K 值的不断增加，各种算法在指标上的变化规律趋于一致。对于准确率，随着 K 值得增加，其呈现先增后减的趋势，验证上文的 K=6 时实验效果最好的结论；随着 K 值的增加，召回率呈现不断增长的趋势，但是增长的幅度由快到慢，逐渐趋于平缓。

本文的 RUTI-CF 算法在准确率值和 F1-measure 指标值上都要优于其他算法。一方面，与 FT-CF 算法对比，虽然 FT-CF 算法也从好友关系和标签两个角度综合考虑用户之间的相似度，但是该算法的准确率值和 F1-measure 指标值均低于 RUTI-CF 算法，说明本文使用融合用户信任和互动影响力考察用户朋友圈中综合推荐信任值更加精确；另一方面，与 FT-CF 算法对比，这个算法仅仅从用户挖掘的标签信息出发，衡量用户的兴趣偏好，且本文算法在找出自己相似用户即"朋友圈"时使用用户偏好模型计算用户基于标签行为的相似度；与 FRSN-CF 算法对比，这种算法利用社交网络中用户的好友交互行为，为用户推荐潜在的好友。但是实验结果表明，本文的 RUTI-CF 算法的准确率值和 F1-measure 指标值都优于这两个算法，说明在社交网络的朋友圈中，仅仅考虑用户的信任关系具有一定的片面性，如果同时考虑用户之间的有向性互动影响力，那么推荐的效果将更为优越。

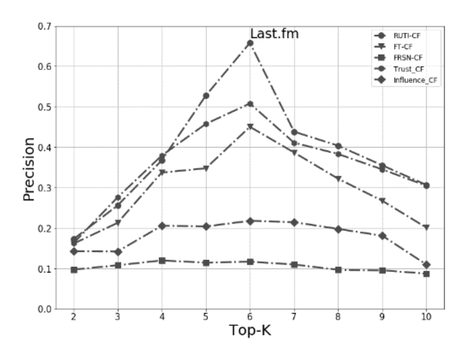


图 4　不同 K 值情况下各算法的准确率

Fig. 4　Accuracy of each algorithm under different K values

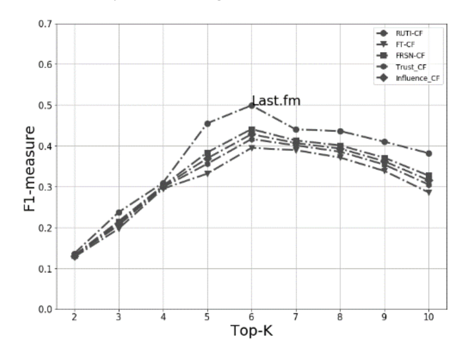


图 5　不同 K 值情况下各算法的 F1-measure 值

Fig. 5　F1-measure value of each algorithm under different K values

5　结束语

从实验结果看，RUTI-CF 算法具有良好的推荐性能，说明在社交平台中，融合在线社交网络用户信任和用户影响力能够更加更加全面地考察用户之间的综合推荐信任度。且本文在衡量用户节点无向性互动影响力，将其与有向性影响系数结合，提出有向性互动影响力，得到的综合推荐信任更加精确。本文推荐模型一方面从直接角度出发考虑用户之间的社交评分信任，另一方面从间接角度通过衡量用户节点的互动影响力对用户节点产生信任关系的可能性的影响程度，使得本实验更符合社交逻辑。因此推荐的准确率和召回率更高。

本文主要讨论静态网络的协同过滤推荐算法，随着系统中用户、项目和标签的数量持续增多，数据稀疏问题等会导致推荐算法的性能降低。结合时序网络的特征，分析总结出在不同的时间窗上用户的相似性特点，从而提出在动态网络中有效的推荐算法模型是本文中进一步研究的重点。

参考文献：

[1] 李良，董宇欣，赵春晖，等. 融合用户信任的协同过滤推荐算法 [J]. 小型微型计算机系统, 2017, 38 (5): 951-955. (Li Liang, Dong Yuxin, Zhao Chunhui, *et al.* Collaborative filtering recommendation algorithm based on user trust [J]. Journal of Chinese Computer Systems, 2017, 38 (5): 951-955.)

[2] 丁小焕，彭甫镕，王琼，等. 融合朋友关系和标签信息的张量分解推荐算法 [J]. 计算机应用, 2015, 35 (7): 1979-1983. (Ding Xiaohuan, Peng Wei, Wang Qiong, *et al.* A tensor decomposition recommendation algorithm combining friend relationship and label information [J]. Journal of Computer Applications, 2015, 35 (7): 1979-1983.)

[3] Zhu Xiaofeng, Zhang Shichao, Jin Zhi, *et al.* Missing value estimation for mixed-attribute data sets [J]. IEEE Trans on Knowledge Data Engineering, 2011, 23 (1): 110-121.

[4] Liben-Nowell D, Kleinberg J. The link prediction problem for social

chinaXiv:201904.00049v1

ChinaXiv合作期刊

networks [C]// Proc of the 12th International Conference on Information and Knowledge Management. New York: ACM Press, 2003: 556-559.

[5] 谢锐，郝志峰，刘波，等. 非对称信息在链接预测中的应用 [J]. 计算机应用, 2018, 38 (6): 1698-1702, 1708. (Xie Rui, Hao Zhifeng, Liu Bo, *et al.* Application of asymmetric information in link prediction [J]. Computer Applications, 2018, 38 (6): 1698-1702, 1708.)

[6] Schall D. Link prediction in directed social networks [J]. Social Network Analysis & Mining, 2014, 4 (1): 157

[7] Bütün E, Kaya M, Alhajj R. Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks [J]. Information Sciences, 2018 (10): 152-165.

[8] 东昱晓，柯庆，吴斌. 基于节点相似性的链接预测 [J]. 计算机科学, 2011, 38 (7): 162-164. (Dong Xiaoxiao, Ke Qing, Wu Bin. Link prediction based on node similarity [J]. Computer Science, 2011, 38 (7): 162-164.)

[9] Aizawa A. An information-theoretic perspective of tf-idf measures [C]// Proc of Information Processing & Management. 2003.

[10] 曲朝阳，宋晨晨，任有学，等. 结合用户活跃度的协同过滤推荐算法 [J]. 东北电力大学学报, 2017, 37 (5): 74-79. (Qu Chaoyang, Song Chenchen, Ren Youxue *et al.* Collaborative filtering recommendation algorithm based on user activity degree [J]. Journal of Northeast Electric Power University, 2017, 37 (5): 74-79.)

[11] 曾安，徐小强. 基于好友关系和标签的混合协同过滤算法 [J]. 计算机科学, 2017, 44 (8): 246-251. (Zeng An, Xu Xiaoqiang. Hybrid collaborative filtering algorithm based on friends relationship and label [J]. Computer Science, 2017, 44 (8): 246-251.)

[12] 吴昊，刘东苏. 社交网络中的好友推荐方法研究 [J]. 现代图书情报技术, 2015, 31 (1): 59-65. (Wu Hao, Liu Dongsu. Research on friend recommendation methods in social networks [J]. Modern Library and Information Technology, 2015, 31 (1): 59-65.)

chinaXiv:201904.00049v1